Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

EREL

Extremal Regions of Extremum Levels

Faraji, Mehdi; Shanbehzadeh, Jamshid; Nasrollahi, Kamal; Moeslund, Thomas B.

Published in:
IEEE International Conference on Image Processing (ICIP), 2015

DOI (link to publication from Publisher):
10.1109/ICIP.2015.7350885

Publication date:
2015

Document Version
Early version, also known as pre-print

Link to publication from Aalborg University

Citation for published version (APA):
Faraji, M., Shanbehzadeh, J., Nasrollahi, K., & Moeslund, T. B. (2015). EREL: Extremal Regions of Extremum Levels. In *IEEE International Conference on Image Processing (ICIP), 2015* (pp. 681-685). IEEE Signal Processing Society. https://doi.org/10.1109/ICIP.2015.7350885

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: May 05, 2023

EREL: EXTREMAL REGIONS OF EXTREMUM LEVELS

Mehdi Faraji, Jamshid Shanbehzadeh

Kharazmi University
Department of Electrical and Computer Engineering
Tehran, Iran

Kamal Nasrollahi, Thomas B. Moeslund

Aalborg University
Visual Analysis of People Laboratory
Aalborg, Denmark

ABSTRACT

Extremal Regions of Extremum Levels (EREL) are regions detected from a set of all extremal regions of an image. Maximally Stable Extremal Regions (MSER) which is a novel affine covariant region detector, detects regions from a same set of extremal regions as well. Although MSER results in regions with almost high repeatability, it is heavily dependent on the union-find approach which is a fairly complicated algorithm, and should be completed sequentially. Furthermore, it detects regions with low repeatability under the blur transformations. The reason for the latter shortcoming is the absence of boundaries information in stability criterion. To tackle these problems we propose to employ prior information about boundaries of regions, which results in a novel region detector algorithm that not only outperforms MSER, but avoids the MSER's rather complicated steps of union-finding. To achieve that, we introduce Maxima of Gradient Magnitudes (MGMs) and use them to find handful of Extremum Levels (ELs). The chosen ELs are then scanned to detect their Extremal Regions (ER). The proposed algorithm which is called Extremal Regions of Extremum Levels (EREL) has been tested on the public benchmark dataset of Mikolajczyk [1]. Our experimental evaluations illustrate that, in many cases EREL achieves higher repeatability scores than MSER even for very low overlap errors.

Index Terms— Maxima of Gradient Magnitude (MGM), Maximally Stable Extremal Region (MSER), Extremal Regions of Extremum Levels (EREL), Feature Detection

1. INTRODUCTION

Feature extraction is an effective stage used at the outset of many computer vision applications including image registration, object recognition, image retrieval, and etc. It consists of a *detection* phase which is followed by a *description* step. Maximally Stable Extremal Regions (MSER) [2] which is a well-known region detection approach, detects interest regions. Given an input image, MSER, which is a level-set based algorithm, thresholds the image with all possible threshold values and detects extremal regions that are most stable. A region is considered stable if its area in a level changes slighter than its area belonging to other levels.

MSER has been deeply studied and widely employed during the last decade. Several authors [3, 4, 5] have proposed improvements on the implementation of MSER. They have concentrated on the algorithmic details of MSER and concurrent construction of the union finding forest and the components' (extremal regions) tree. On the other hand, there are lots of extensions over the original version of MSER, as those in [6, 7, 8, 9, 10, 11, 12]. In [9], an integrated algorithm, ED-MSER, is proposed which combines MSER [2] with SIFT [13] and a filtering strategy. The notion of enclosed regions detected by setting several thresholds was introduced in [10]. Enclosed regions consist of External Enclosing Regions (EER) and Internal Enclosed Regions (IER). Recently, MSER has been adapted to work with scale-space theory in [14, 11, 12]. Forseen and Lowe [14] introduced a multi-resolution version of MSER and used it to construct a descriptor for detected extremal regions. In [12], MSER is employed on the Difference of Gaussian (DoG). The stable regions are selected if the barycenter of a region is surrounded by at least ten barycenters of other regions in its adjacent scales.

As illustrated in [1], MSER obtained low repeatability when blur transformations happen. This can be mostly because of its stability criterion. The stability criterion considers mainly the area of the regions as a crucial parameter. So, the stable regions are those regions with the least change during a range of thresholds which considers no information about the boundaries of the regions during the detection of stable regions. Particularly, blur transformation manipulate the boundaries of the regions, so the area-based stability criterion of MSER, when the image transformed by blurring, can result in low repeatability. To deal with this problem Kimmel [15] proposed other stability criteria based on the length of the boundaries. It can be seen that various types of transformations may need different kind of stability criterion to achieve ideal performance. Another issue with MSER is that it mostly include a rather complicated step of enumeration followed by a cleaning up step. The proposed system in this paper presents a *completely novel region detection algorithm* in which the above-mentioned issues are tackled by includ-

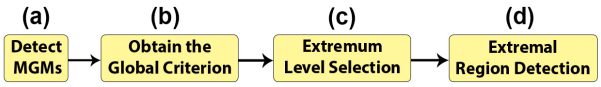

Fig. 1: The block diagram of the proposed system.

ing the information of the region boundaries. To do so, we introduce a kind of interest points, *Maxima of Gradient Magnitudes* (MGMs), which are mostly concentrated around the edges (region boundaries). Therefore, they are used as prior information for detecting invariant regions in our region detection algorithm. The experimental results on a challenging benchmark dataset show superiority of the proposed algorithm over the state-of-the-art region detection algorithms. The rest of the paper is as follows. The details of the proposed algorithm and the introduced MGM points are explained in the next section. The experimental results are given in Section 3, and finally the paper is concluded in Section 4.

2. THE PROPOSED METHOD

The block diagram of our novel proposed algorithm is shown in Fig. 1. We first detect MGM points (Fig. 1a). Then, we binarize the image by applying all possible thresholds on it and obtain a Global Criterion (GC) based on the information provided by both white pixels of the thresholded image and the MGMs (Fig. 1b). After that, we select a number of Extremum Levels (ELs) based on the GC vector (Fig. 1c). In the last step (Fig. 1d) we detect extremal regions in only those selected ELs that intersect with at least one MGM. The above steps are explained in the following subsections, respectively.

2.1. Maxima of Gradient Magnitude (MGM)

Given an input image, I, we first obtain its gradient, ∇I, by a simple gradient filter like Soble. An MGM is a point $p(x, y)$ in ∇I that has two conditions: first, it has a maximum value of the gradient magnitudes among their local neighborhood points with radius r, as:

$$\|\nabla I(p)\| > \|\nabla I(P)\|, \forall P \in N(p, r) \quad (1)$$

where $N(p, r) = \{P | \forall P \in I, \|P - p\| < r\}$ is the neighbor function. Second, the mean of the gradient magnitudes of its neighbor points, should be larger than a threshold:

$$E[\|\nabla I(N(p, r))\|] \geq \alpha . E[\|\nabla I\|] \quad (2)$$

where $E()$ is a mean function, and α is an arbitrary coefficient which controls the number of resulted points. The chosen threshold, $\alpha . E[\|\nabla I\|]$, is a function of the mean of the gradient magnitudes image, because according to [16], such a function can keep a fair amount of high informative points along edges. Checking the above two conditions for all the pixels of the input image, a binary image, M, can be generated in which the positions of the MGMs are highlighted.

Fig. 2: MGM^+s and MGM^-s, shown by + and -, respectively, for a part of the "bikes" images of [1].

2.1.1. MGM^+ and MGM^-

Two sets of regions, Q^+ and Q^-, can be detected from an input image when any type of level set methods, like the one used in this paper, is employed:

- The first set, Q^+, contains regions that evolve from brighter surfaces to darker boundaries. These regions can be detected from the original image by thresholding at different levels. Each of these thresholds result in a binary image, $T^+_\vartheta(x, y)$, as:

$$T^+_\vartheta(x, y) = \begin{cases} 1 & \text{if } I(x, y) \geq \vartheta \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

  where ϑ is the current threshold. The range of the thresholds obviously depends on the number of the bits used per pixel.

- The second set, Q^-, contains those regions that evolve from darker surfaces to brighter boundaries. To detect these regions we use:

$$T^-_\vartheta(x, y) = \begin{cases} 1 & \text{if } I(x, y) \leq \vartheta \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Since region boundaries of each of the above two sets have different characteristics, we divide the image containing the detected MGMs of the previous step (section 2.1), M, into two sets, namely MGM^+ and MGM^-, each of which contains points that can be used in its corresponding evolving technique. To achieve this, an extremum level (ϑ_x) needs to be found as a division criterion.

How can such a ϑ_x be almost pointed out? We propose to use the median of the distribution of MGMs for this purpose. This is because such a value (median of the distribution) can divide the function into two parts. The total summation of the probability values of one part ($^+$) is sufficiently close to the total summation of the other part ($^-$). So, it can fairly divide MGMs. To calculate the median, we first need the intensity value of the MGM points, which can be obtained by element-wise multiplication of M and the input image, I. Then P_ϑ which is the normalized histogram of MGMs at level ϑ can be easily calculated. The median of the distribution (ϑ_x) is a level where $\sum_{j=0}^{\vartheta_x} P_j >= \frac{1}{2}$ and $\sum_{j=0}^{\vartheta_x - 1} P_j < \frac{1}{2}$. This obtained median, ϑ_x, is then used to threshold the image M containing the MGMs and separate the appropriate interest

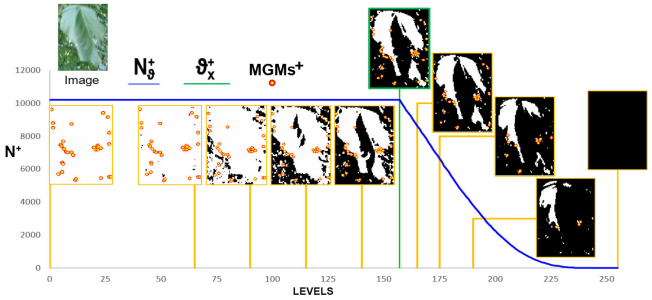

Fig. 3: Regions and MGMs intersections in the levels evolution of the image, on x and y axis are $N+$ and the levels, respectively.

point for each types of extremal regions ($^+$ and $^-$), as:

$$M^+(x,y) = \begin{cases} 1 & \text{if } M(x,y) \cdot I(x,y) \geq \vartheta_x \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

A similar equation can be considered for obtaining M^- as:

$$M^-(x,y) = \begin{cases} 1 & \text{if } M(x,y) \cdot I(x,y) \geq \Theta - \vartheta_x \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where Θ is the maximum number of thresholds (gray-levels).

Fig. 2 shows two types of MGMs extracted from a same part of the "bikes" image. It can be seen that detected $M^+(x,y)$ are located in the regions with a brighter surface and a darker boundary and $M^-(x,y)$ are located in the regions with a darker surface and a brighter boundary.

2.2. Global Criterion (GC)

Following the block diagram of the algorithm in Fig. 1, having found the MGMs, the next step is to obtain a GC. To do so, the number of intersection points between the binary image of level ϑ and M^+, is obtained, which is denoted by N^+_ϑ:

$$N^+_\vartheta = \sum_{x=1}^{X} \sum_{y=1}^{Y} T^+_\vartheta(x,y) \cdot M^+(x,y) \quad (7)$$

For each type of regions ($^+$ and $^-$) we should perform the following similar steps, however we only explain the steps belong to $^+$ type. The reader can do the same process for the $^-$ type of regions.

To find levels of significant variations, firstly vectors V^+ should be calculated. Each vector is defined as a ratio of concurrent change of the total number of the white pixels in the thresholded image and N^+_ϑ (see Fig. 3). To do so, we define V^+ using N^+_ϑ as:

$$V^+_\vartheta = \frac{N^+_\vartheta}{1 + \sum_{x=1}^{X} \sum_{y=1}^{Y} T^+_\vartheta(x,y)}, 0 \leq \vartheta \leq \Theta \quad (8)$$

The underlying variations of V^+ are suitable clues for indicating levels in which regions change remarkably. Finally, using central difference, the GC, Ψ, which is the weighted second order derivative of vector V, is defined as:

$$\Psi^+_\vartheta = \frac{\frac{d}{d\vartheta} V^+_{\vartheta-1} - \frac{d}{d\vartheta} V^+_{\vartheta+1}}{1 + \frac{d}{d\vartheta} V^+_\vartheta} \quad (9)$$

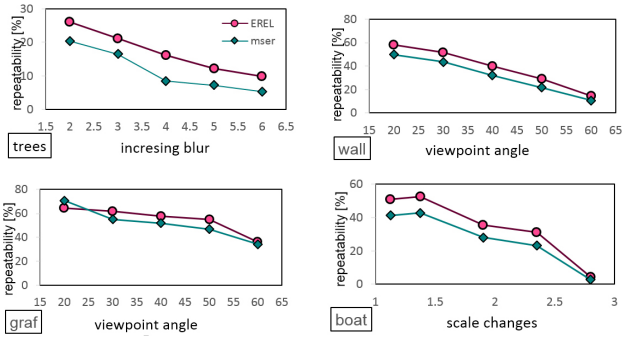

Fig. 4: Repeatability scores achieved by 20% overlap errors for image sets of dataset [1].

2.3. Extremum Levels (ELs) Selection

Following the block diagram of the algorithm in Fig. 1, having obtained the Ψ^+_ϑ, we need to find ELs$^+$. A level like ϑ belongs to the set of EL$^+$ if its Ψ^+_ϑ, is a local minimum or maximum. To select EL$^+$, each cell of Ψ^+_ϑ is hence compared with its δ previous cells and δ subsequent cells. δ shows the radius of the neighborhood window and represents the number of adjacent levels which are involved in the process of local minima selection (Fig. 1).

2.4. Extremal Regions Detection

Extremal regions can finally be detected from each elements of EL$^+$ and EL$^-$ by any arbitrary connected component analysis algorithms or a labeling strategy. Note that for each vector (EL$^+$ and EL$^-$), we run the algorithm separately. So, the following explanation should also be considered for EL$^-$. The algorithm starts from the first indicated EL$^+$, extracts the extremal regions of that level and chooses only those extremal regions that intersect with at least one MGM. After that, the MGMs belonging to the selected extremal regions will be ignored in the next ELs$^+$. This process continues till all elements of ELs$^+$ be processed. If a region does not intersect with MGMs, we can imply that the region is not stable enough. On the other hand, an MGM might intersect with no region. This shows that the point has been detected wrongly because of the presence of noise. So, both of these conditions complement each others. In addition, MGMs actually help both selecting regions and cleaning up the unwanted extremal regions. However, it should be noted that no direct cleaning up nor clustering and enumeration is performed by EREL. Since, our proposed method detects extremal regions belonging to both ELs$^+$ and ELs$^-$, we call it: Extremal Regions of Extremum Levels (EREL).

3. EXPERIMENTAL EVALUATION

To demonstrate the superiority of the proposed method against the competing state-of-the-art method of MSER, a common criterion has been used, namely *repeatability* [1].

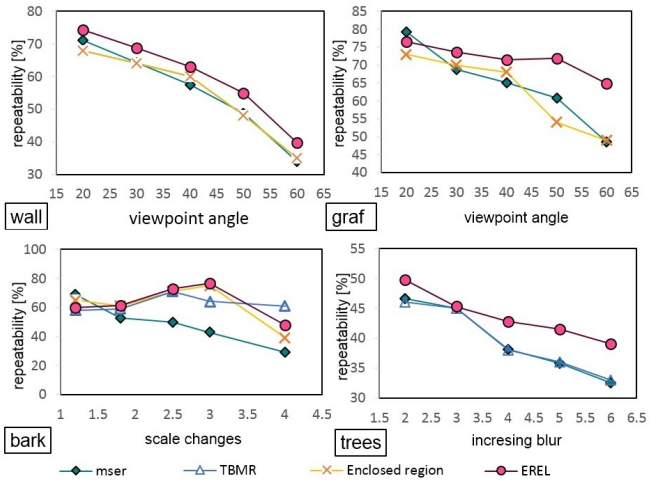

Fig. 5: Comparing the Repeatability scores of EREL with MSER and two very recent methods of TBMR [17] and Enclosed region [10] (Overlap error is 40%). Please note that some of the plots are overlapping.

Using this criterion, we have evaluated the proposed algorithm on the image sequences of the widely known benchmark dataset of Mikolajczyk, affine covariant dataset [1]. The images in this dataset have gone through different image degradations, including blur (by "tress" sequences), viewpoint (by "graf" and "wall" sequences), scaling and rotation (by "boat" sequence). The repeatability score provides a quantity value of performance including the accuracy of localization and is defined as the ratio between the number of region-to-region correspondences and the smaller number of regions detected in one of the images [1]. The performance of the EREL, its repeatability scores of four image sets of the dataset are shown in Fig. 4 with an overlap error of 20%. It can be seen that in several cases, especially in the textured sequence with blur transformation ("trees"), the proposed system outperforms MSER. However, the degree of the improvement changes from one image to another, since the contents of the images in the dataset are very different. For example, for the "trees" and "wall" images, the EREL results in much higher repeatable regions, compared to the other images. This is because, these images represent textured type of scenes producing greater number of MGMs. On the other hand, the structure type scenes contains very strong edges. This means that the MGMs can be better detected on the boundaries of shapes, which at the end results in better performance of the system.

Moreover, we compare the repeatability scores of the proposed EREL method against the very recent extensions of MSER which have reported their results on the benchmark dataset of [1]. These systems are TBMR [17] and Enclosed region [10]. The results of the comparison (for an overlap error of 40%) are shown in Fig. 5. It can be seen from this figure that for both structure type scenes ("graf" and "bark") and texture like images ("wall" and "trees") EREL outperforms the other methods.

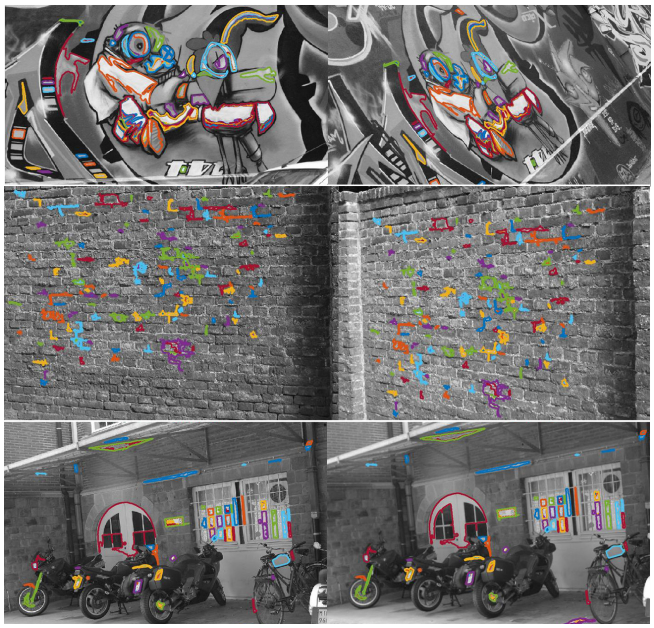

Fig. 6: The corresponding regions found by the proposed EREL method in three pairs of images from [1] dataset. In each row (pair) the left image is the reference image and the right image is the transformed image (Overlap error = 10%).

Finally, Fig. 6 shows the output of the system for three image pairs from the Mikolajczyk dataset of [1].

4. CONCLUSION

In this paper, we have proposed a novel method for extremal region detection. The most notable Maximally Stable Extremal Regions (MSER) was the inspiration of our proposed algorithms. It detects stable repeatable regions by using a union find structure. Although MSER is efficient, it detects low repeatable regions when image is transformed by blurring. This is a cause of not using any information about the boundaries of the regions. We have shown in this paper that including such information in the process of finding the maximally stable regions not only removes the need for the cumbersome step of regions enumerations and the cleaning up step of MSER, but also results in a region detector that outperforms MSER. This has been proven through experimental results on the popular benchmark dataset of Mikolajczyks in [1] which imposes different image degradations, such as blur, viewpoint change, and etc, to its image sequences. The experimental results show that the proposed algorithm outperforms MSER in detecting more accurate repeatable regions. We are planning to extend the proposed algorithm to video sequences and utilize temporal information in our future work.

5. REFERENCES

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vision*, pp. 43–72, 2005.

[2] J Matas, O Chum, M Urban, and T Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, pp. 761 – 767, 2004.

[3] Erik Murphy-Chutorian and Mohan M Trivedi, "N-tree disjoint-set forests for maximally stable extremal regions.," in *BMVC*, 2006, pp. 739–748.

[4] F. Kristensen and W.J. MacLean, "Real-time extraction of maximally stable extremal regions on an fpga," in *IEEE Int. Symp. on Circuits and Systems,*, 2007, pp. 165–168.

[5] David Nistr and Henrik Stewnius, "Linear time maximally stable extremal regions," in *ECCV*, pp. 183–196. 2008.

[6] Friedrich Fraundorfer, Martin Winter, and Horst Bischof, "Mscc: Maximally stable corner clusters," in *Image Analysis*, Lecture Notes in Computer Science, pp. 45–54. 2005.

[7] P.-E. Forssen, "Maximally stable colour regions for recognition and matching," in *CVPR*, 2007, pp. 1–8.

[8] M. Perdoch, J. Matas, and S. Obdrzalek, "Stable affine frames on isophotes," in *ICCV*, 2007, pp. 1–8.

[9] Liang Cheng, Jianya Gong, Xiaoxia Yang, Chong Fan, and Peng Han, "Robust affine invariant feature extraction for image matching," *Geoscience and Remote Sensing Letters, IEEE*, pp. 246–250, 2008.

[10] Wei Zhang, Q.M. Jonathan Wu, Guanghui Wang, Xinge You, and Yongfang Wang, "Image matching using enclosed region detector," *Journal of Visual Communication and Image Representation*, pp. 271 – 282, 2010.

[11] Luo Ronghua and Min Huaqing, "Multi-scale maximally stable extremal regions for object recognition," in *IEEE Int. Conf. on Information and Automation*, 2010, pp. 1799–1803.

[12] Meng-meng Zhang, Ze-ming Li, Hui-hui Bai, and Yan Sun, "Robust image salient regional extraction and matching based on dogss-msers," *Optik-International Journal for Light and Electron Optics*, pp. 1469–1473, 2014.

[13] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, pp. 91–110, 2004.

[14] P.-E. Forssen and D.G. Lowe, "Shape descriptors for maximally stable extremal regions," in *ICCV*, 2007, pp. 1–8.

[15] Ron Kimmel, Cuiping Zhang, Alexander M Bronstein, and Michael M Bronstein, "Are mser features really interesting?," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 11, pp. 2316–2320, 2011.

[16] F Bellavia, D Tegolo, and C Valenti, "Improving harris corner selection strategy," *IET Computer Vision*, vol. 5, no. 2, pp. 87–96, 2011.

[17] Y. Xu, P. Monasse, T. Geraud, and L. Najman, "Tree-based morse regions: A topological approach to local feature detection," *IEEE TIP*, pp. 1–1, 2014.